DOCUMENT RESUME

ED 383 298 IR 017 153

AUTHOR Hale, Michael E.; And Others
TITLE The Portable Usability Testing Lab: A Flexible Research Tool.
PUB DATE 95
NOTE 10p.; In: Proceedings of the 1995 National Convention of the Association for Educational Communications and Technology (AECT) (17th, Anaheim, CA, (17th, Anaheim, CA, 1995); see IR 017 139.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; Computer Software; *Educational Research; Educational Technology; *Evaluation Methods; Higher Education; *Laboratories; Online Systems; Performance; *Research and Development
IDENTIFIERS Computer Users; Interactive Computer Systems; Learning Environment; *Multimedia Technology; *Performance Support Systems; University of Georgia; Video Technology

ABSTRACT

A group of faculty at the University of Georgia obtained funding for a research and development facility called the Learning and Performance Support Laboratory (LPSL). One of the LPSL's primary needs was obtaining a portable usability lab for software testing, so the facility obtained the "Luggage Lab 2000." The lab is transportable to any site where interactive multimedia or any other type of software is being used for education, training, information, or performance support purposes. It includes a remote-controlled video camera that can be focused the record the user's face, work on a desk, the user's computer, keyboard and mouse, or any other aspect of the user environment. The system simultaneously records whatever appears on the user's screen. Researchers can control what is recorded (e.g., most of the user's screen along with a small insert of the user's facial expressions or body language). The portable usability lab was used for two purposes: (1) as part of a formative evaluation of a computer-based learning environment for statistics, called "StatSim"; and (2) to examine the types of interactions initiated by learners who were in two different experimental groups using an intelligent tutoring system, "MSRT Tutor." The research data yielded by the portable usability lab provides an improved basis for guiding the design and implementation of instructional technology. The lab is pictured in one figure. (Contains 13 references.) (MAS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 383 298

Title:

The Portable Usability Testing Lab: A Flexible Research Tool

Authors:

Michael E. Hale, Ph.D.
Michael A. Orey, Ed.D.
Thomas C. Reeves, Ph.D.

Learning and Performance Support Laboratory (LPSL)
College of Education
The University of Georgia
611 Aderhold Hall
Athens, GA 30602-7144

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. ZENOR
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR017153

BEST COPY AVAILABLE

Introduction

In 1992, the Georgia Research Alliance released a "Request for Proposals" (RFP) for establishing an infrastructure for research and development focused on advanced telecommunication technologies and applications in the state of Georgia. A group of faculty (including the authors of this paper) in the College of Education at The University of Georgia responded to this RFP and obtained funding for a new R & D facility called the Learning and Performance Support Laboratory (LPSL). The four areas of emphasis in the R & D efforts of the LPSL are 1) interactive learning environments, 2) electronic performance support systems, 3) alternative performance assessment systems, and 4) information access systems.

The LPSL is committed to knowledge acquisition, theory construction, and the highest ideals of applied research and development, especially as they relate to education, training, and performance problems that must be solved to ensure the economic viability of state of Georgia. Advanced telecommunications, along with genetic engineering and environmental technologies, have been identified as the three key industries for Georgia in the 21st Century. The LPSL is committed to partnerships and collaborations with other institutions, businesses, and agencies in the state. Current partners include R & D labs and centers at the five institutions (Georgia Institute of Technology, Georgia State University, Emory University, Clark-Atlanta University, and the Medical College of Georgia) which together with The University of Georgia (UGA) form the Georgia Research Alliance (GRA). Equally important partners are other Georgia public institutions (e.g., colleges, schools, technical training centers, etc.), businesses and industries (e.g. AT&T, BellSouth, and CNN), and government agencies (e.g., the U. S. Army Signal School at Ft. Gordon). The LPSL is also establishing substantive collaborations with partners from around the USA and the rest of the world. For example, the LPSL has established a relationship with the "Cooperative Multimedia Centre" (CMC) in Perth, a collaboration involving four universities in Western Australia.

Usability Testing

To achieve its R & D goals, faculty associated with the LPSL realized that they would need state-of-the-art research equipment. One of the primary needs was obtaining a portable usability lab for software testing. According to Shneiderman (1987), usability is a combination of the following user-oriented characteristics: 1) ease of learning, 2) high speed of user task performance, 3) low user error rate, 4) subjective user satisfaction, and 5) user retention over time. Hix and Hartson (1993) and Nielsen (1993) provide guidance to evaluating user interface issues, a process known as usability testing. Usability testing is especially critical in the design, dissemination, and implementation of interactive multimedia for education, training, performance support, and information access (cf., Blattner & Dannenberg, 1992; Laurel, 1990; Polson, 1988, Shneiderman, 1987).

A proposal was written and funding for the portable usability lab was obtained in 1993. Usability Systems, Inc., an Atlanta based company that builds usability labs for testing software, responded to our request for bids for a portable system with the "Luggage Lab 2000" (see Figure 1). This lab is transportable to any site where interactive multimedia or any other type of software is being used for education, training, information, or performance support purposes. The "lab" includes a remote-controlled video camera that can be focused on the user's face, work on a desk, the user's computer, keyboard and mouse, or any other aspect of the user environment considered important in the study. The system simultaneously records the user's computer screen. Researchers sit at a control panel that allows them to observe the user(s) directly or on any of the video screens displaying selected aspects of the context. Researchers can control what is recorded, e.g., most of the user's screen along with a small insert image of the user's facial expressions or body language.

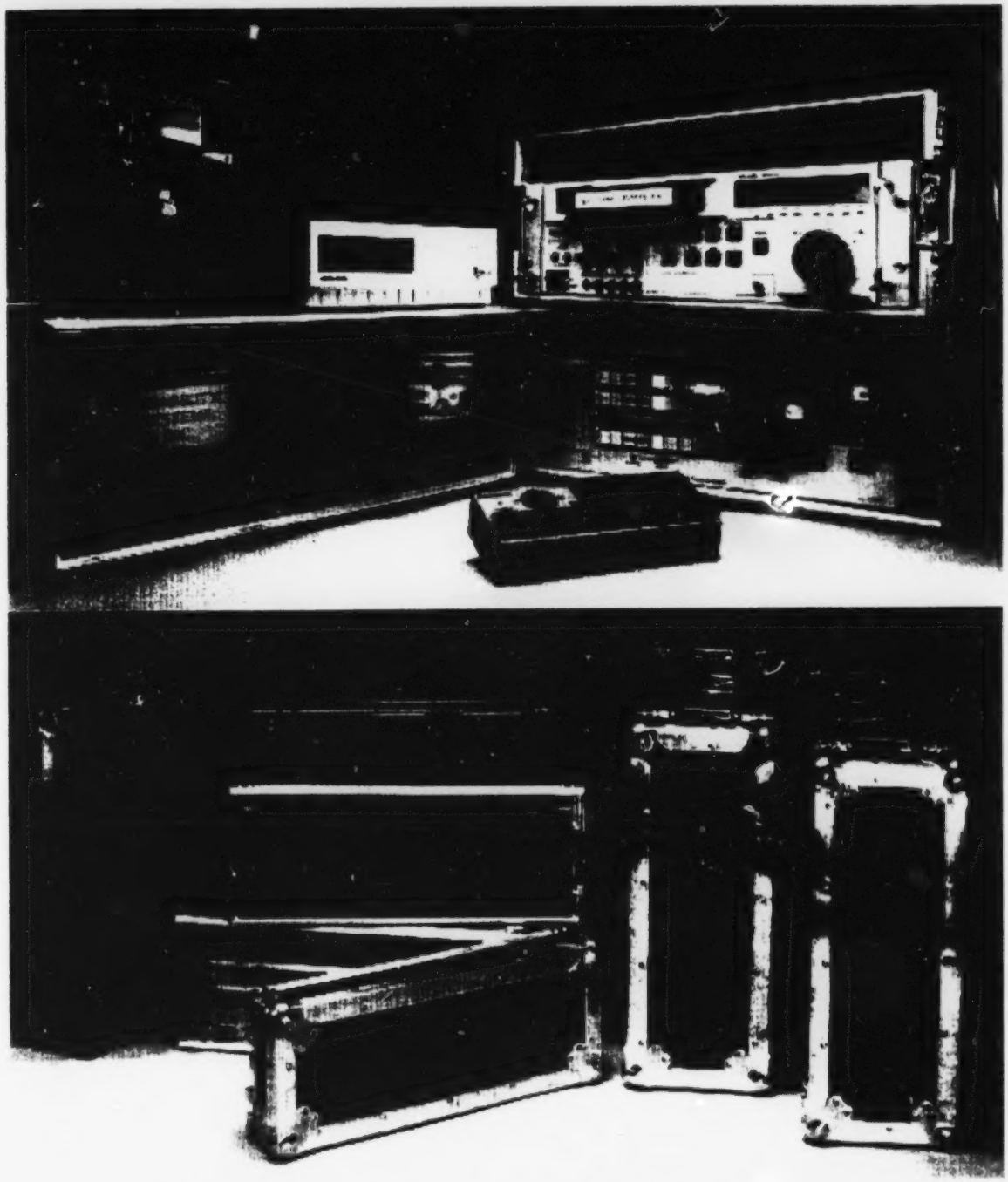

Figure 1. Portable usability lab. (FMI, contact Usability Systems, Inc., 404/475-3505.)

BEST COPY AVAILABLE

Commercial software developers have employed "usability labs" for formative evaluation of software applications for many years (Gomoll, 1990). For example, both AT&T and IBM maintain fixed usability testing labs in the Atlanta area. These fixed usability labs generally consist of two rooms separated by a one-way glass window (see Figure 2 below). In one room, a computer user sits at a desk and uses the application being evaluated, e.g., a new spreadsheet program. Several video cameras mounted in the room are focused on various aspects in the room. In the other room, evaluators and designers sit at control panels where they can simultaneously observe the user in the room through the one-way glass or any of the video screens displaying selected aspects. The user may be instructed to "think aloud" as he/she uses the program, e.g., talk about why certain choices are made or describe any confusion about the program's interface. Alternatively, the evaluators may question the user via headsets or speakers about why he/she has done certain actions. Typically, these sessions are videotaped for later analysis and documentation. Some fixed usability labs feature a third room where clients can observe the usability testing as it is being conducted.

Figure 2. Fixed usability laboratory (artwork by Lih-Juan Chanlin).

The portable software usability lab is patterned after these commercial labels, but rather than forcing users to come to a lab and test software in an artificial environment, the portable lab allows the users to stay in their own environment. We believe that this increases the validity of both software testing and research studies. This paper is intended to describe our procedures for using this research tool, present some of the preliminary results we have obtained with it, and suggest recommendations for further research.

Procedures for Usability Testing

The portable usability lab enables researchers to collect both quantitative and qualitative data related to issues such as user interface, mental models, navigation, documentation utility, effectiveness, and efficiency. A variety of research and evaluation protocols are possible using this type of tool (Hix & Hartson, 1993; Nielsen, 1993).

Our usability lab comes in four large cases that stack onto a dolly for rolling (see Figure 1). Detailed instructions describe step by step procedures for connecting system components together. All cables are color coded to assist in setup. Beyond the normal components of the Luggage Lab 2000, Usability Systems, Inc. added picture in a picture capability to the lab. This allows us to record both the full computer screen and an inset of the user. An FM microphone is provided for think aloud and interview procedures. The camera has remote control features to pan and tilt for added flexibility in recording the interactions of the user.

This flexible system can be used to support and enhance many methods of usability testing. Nielsen (1993) identified the following methods for gathering usability data: observation, think aloud, questionnaires, interviews, focus groups, logging actual use and user feedback. Researchers and developers should select the appropriate methods to collect data to address different usability issues and questions.

Each of these methods has different strengths and weaknesses, and combining different methods is often necessary to improve overall usability testing. For example, you might want to ask questions to users during observations, but asking questions during an observation can change what the user would naturally do. One solution is to record the user with the portable usability lab, and later play the tape back to the user and ask questions. The tape assists the user in recalling the recorded session. In addition the same tape can be shown to human factors experts for their advice and interpretations. A focus group can review videotapes of users in their actual working conditions to stimulate discussions.

Both research projects and software evaluations can involve the use of experts to judge the performance of users on various types of tasks. Reliability is an important issue whenever human judges are used. Having videotaped data scored by multiple experts can provide reliability information about the data collection process. Collecting data about benchmark tasks is another use of the portable usability lab. A benchmark task is a common activity the user performs with the technology. These benchmarks are selected by the developer to measure quantitatively the interface design (Hix & Hartson, 1993). The usability system can record the user's performance on benchmarks for later analysis.

In addition to these planned benchmark tests, you may observe users with the portable usability system over time in their natural environment doing what they select to do. This is more like naturalistic research than benchmark evaluation. Benchmarks are focused and efficient whereas natural observation is time consuming and less directed. However, the generalizability of findings from naturalistic research may be greater. Both are valuable and can be supported with the portable usability system.

Many times in designing an interface you may have multiple options for designing an interaction with the computer. Local "rapid prototyping" is the creation of multiple designs for small components of your software (Hix & Hartson, 1993). By comparing tasks performed with each option of interacting, you can feel more confident about which design to use.

An example of multiple designs for a computer interaction was considered by the first author in the design of an angle measurement tool, a computerized protractor. One way to measure an angle is to click on three separate points that form an angle (one on the first leg, one at the vertex, and one on the other leg of the angle). Another way to design the protractor is to click on the vertex and drag along one side and release, then you push or pull to create a spanning angle that reaches the other leg (Hale, Gustafson & Yeany, 1989). The first interaction is more straight forward. The latter has the advantage of positive and negative values depending visually on the direction of spanning. By creating both versions in a small prototype, you may test these for inclusion into the final product.

Finding appropriate users who will allow you to videotape them at work with a prototype system is sometimes difficult. Experts suggest only using a few participants that match your target users (Hix & Hartson, 1993). You should supplement these by using one or two experts in human-computer interface design to review your software. With so few people interacting with your software, you may want to maximize the data you have by doing detailed analysis using a tool such as Videotape Analyzer (Tsao, Hale & Fan, 1994). This software tool supports repeated detailed qualitative analysis of video. The portable usability lab is ideal for producing videotapes that can be analyzed with this tool.

Results of Usability Testing

The usability lab at UGA was used in late 1994 for two different purposes. i.e., formative evaluation and educational research. The first application was part of a formative evaluation of a computer-based learning environment for statistics called *StatSim*. The second was to examine the types of interactions initiated by learners who were in two different experimental groups using an intelligent tutoring system called the *MSRT Tutor*. The best way to understand the power of the usability lab is to describe our experiences with these two applications.

In the case of the *StatSim* evaluation, the usability lab was used in the College of Education at UGA. Therefore, the portability part of the system was not utilized to its fullest extent. The system was setup in a small limited access hallway that was adjacent to the room where the software testing was to take place. This was quite useful in that the people who were being recorded did not actually see any of the equipment

that was being used. We believe that this would have resulted in a heightened anxiety level within the people who were learning from *StatSim*. However, because there was a cinder block wall between the usability equipment and the learner, the audio transmit/receive ability was problematic. Later uses of the lab attended to this placement problem by making sure that there were no obstructions between the remote microphone of the user and the receiver built into the usability equipment. Office partitions block the view, but do not block the radio signal. Another, as yet unexplored possibility, would be to add to the lab a remote antenna that can be placed in the room where the user is located.

The intent of the *StatSim* evaluation was to improve the software. The program had been developed by Liu Zhang, a graduate student in the Artificial Intelligence Program at UGA. This program was part of his requirements for a Master's degree. Within the program, he used elements of constructivist learning theory to allow the learner to freely explore problems in statistics. However, once the learner navigated to the point where problem solving was taking place, the system was then able to model the learner's problem solving and provide appropriate feedback.

The *StatSim* program was written in Visual Basic. Many of the changes that were made to the software were the result of learners' expressed lack of understanding of particular icons, text messages, or information windows. While it would be beyond the requirements of this paper or interest of the reader to document all the changes here, it should be noted that confusing items that were presented to the learner as they interacted with the *StatSim* program were located not only by their verbal comments, but also by their facial expressions as they were being shown different aspects of the program. The videos captured by the usability lab enabled us to locate and modify these confusing items.

Curiously, during the interactions, *StatSim* caused the computer to "lock up" on a seemingly unpredictable basis. Each time this would happen, the computer was restarted and the learner continued. At first, the developer was at a loss as to why the system was crashing. Upon review of the video tape that was created by the usability lab, it was discovered that the system would crash on the fifth time that the learner would access the calculator (a resource program that comes with Microsoft Windows). Apparently, each time the learner accessed the calculator, when they were done, they would click back on a *StatSim* screen. The calculator program was therefore still running. The next time the learner clicked on the calculator, a new instance of the calculator was created. Finally, on the fifth access to the calculator, the memory was exceeded and the computer would "lock up". The video obtained with the usability lab enabled the developer to track down this peculiar problem.

A different application of the usability lab took place when we used it during an experiment with the *MSRT Tutor*. The tutor was originally developed by Michael Orey to determine if an interesting ITS like system could be built to run on a 286-based MS-DOS computer system. It has since become a primary tool for teaching soldiers how to operate the MSRT radio. It has been used by over 10,000 soldiers for this purpose. Some of these soldiers use the program as part of their formal classroom training, others use the system in the field as is needed. The experiment examined various sustainment schedules in order to help soldiers maintain procedural knowledge in memory. One group used the Tutor every two weeks, another group every three weeks, and the last group every six weeks. Not surprisingly, the two week group was better than the three week group that was better than the six week group. The question that confronted us was: What kinds of things were learners doing on the Tutor in each of these three groups? We used the usability lab to document the interactions that the learners made on the system. We thought that the differences observed would help us understand or explain the learning differences that were measured between the groups. Two learners from each group were randomly selected to use the computer that was hooked up to the usability lab. A detailed analysis was performed on each of the video tapes that were created while these learners interacted with the Tutor.

The portability of the usability lab was very helpful in this application. This experiment took place at Fort Gordon, Georgia, a ninety-minute drive from UGA. The usability lab was loaded into a small car (Ford Escort) driven by the second author of this paper. The entire system breaks down into four large suitcases. One was placed in the trunk, two in the back seat and one in the front passenger seat. A classroom at Fort Gordon was used for data collection. The computer was setup in the front of the room facing forward and the usability lab was behind the learner. We did not want to make the learner feel too much anxiety while we collected the data, but this setting required us to place the lab in the same room as the learner. On the

other hand, this configuration did not have the same transmit/receive problems that we had experienced with the formative evaluation in the College of Education.

The results of the Tutor analysis resulted in two important findings. First, the six and three week groups performed many more interaction patterns, an indication that the learners were forgetting how to use the program (the six week people more so than the three week people). This type of interaction would not result in better performance on the actual task. In fact, knowing how to use the Tutor per se is of little benefit to performing the task.

The second type of interaction that was different between groups requires a little more explanation of the treatments. The two week group performed the tasks twice every two weeks for six weeks (a total of 2 X 3 or six times). The three week group performed the tasks three times every three weeks for six weeks (a total of 3 X 2 or six times). The six week group performed the task six times once on the sixth week following the initial training. In this way, we could vary the schedule while holding the number of repetitions constant. The result of this design as shown in the videos produced by the usability lab was that there were diminishing benefits on the repetitions as the number increased. The sixth time the six week group performed the task was largely a non-memory intensive effort on the part of the learner (what youth today call a "no brainer"). According to this data, one might conclude that if we were to separate the repetitions by a short break for the six week group, their effort at the task might improve and so would their performance.

This study also revealed statistically significant differences in outcomes between soldiers who performed a fixed number of problems and those that worked towards a criterion level of performance during each of their three sustainment episodes. It was difficult to understand why we would get statistical differences between the groups until we went back and examined the data on the video tapes to compare learners in each group. As it turns out, we only had data on two of subjects from the Fixed condition and only one subject from the Criterion condition. The results of our analysis are presented in Table 1. There were three primary issues or categories of behaviors that were observed during their interaction. These issues clearly demonstrate that the Criterion One person is focusing his efforts on meeting the criteria imposed by his version of the tutor. If this pattern can be generalized to the entire group, then a possible explanation for why there was a performance difference is that the criterion group appears to be focusing their attention and efforts towards meeting the criteria while the Fixed group is focusing their attention and effort more directly on learning the content of the Tutor. If this assumption is true, it is clear why we found this performance difference and it is clear that having learners perform a fixed number of trials is a preferred teaching strategy for sustainment learning.

Table 1.
Comparison between two subjects in the Fixed group and one subject in the Criterion group

Issue	Fixed One	Fixed Two	Criterion One
View of Coach	Ask if Click Coach OK; Uses Coach frequently; Says he, "needs Coach"	Says he already knows about MSRT; Feels he doesn't need Coach; Only uses Coach once	Avoids Coach; Recognizes that Coach will prevent progress
Learning Goals	Finishes the 2 passes and expresses relief; Doesn't want to do it again	Tries to finish within 5 minutes	Is totally focused on meeting the criterion for completing the Tutor
Expressions of Confidence	Creates many excuses for errors	"I'm surprised I remember"	Anticipating performance feedback, "Do I have any errors?"

While it is difficult to fully capture the entirety of the impact of the usability lab in these short anecdotes, we hope that the reader begins to understand the potential benefit of this type of research tool. We have several new research projects underway where the usability lab will have great benefit. For example, one of our doctoral students is focusing her research on the mental models users construct of an on-line public access catalog (OPAC) (Kelly, 1994). She is using the portable usability lab to record users as they

perform searches on GALIN (Georgia Library Information Network) in the main library where she works as a reference librarian. She then reviews each video with a user, asking them to explain why they made certain choices. Research on mental models is an important direction in the study of user interfaces for interactive multimedia and other applications of instructional technology (Jih & Reeves, 1992; Leiser, 1992).

Conclusion

The importance of the "usability" approach to research and evaluation is considerable (Nielsen, 1993). Currently, instructional designers and researchers have an inadequate base of knowledge about how users react to and learn with multimedia programs and other types of computer programs such as electronic performance support systems (Gery, 1991). The research data yielded by the portable usability lab provides an improved basis for guiding the design and implementation of instructional technology systems. For example, we believe that the enhancement of our understanding of interactive multimedia user interfaces can improve the dissemination, implementation, and effects of using multimedia at all levels of education and training. Initial marketing of multimedia has succeeded primarily on the basis of selling the "bells and whistles" of the technology, but now school boards, superintendents, parents, and taxpayers are beginning to demand research-based evidence that multimedia enhances learning. Fundamental understanding of how interactive multimedia is used by teachers, students, trainees, etc. is an essential part of that evidence. The portable usability testing lab provides us with precisely that kind of evidence.

Background of Authors

MICHAEL E. HALE, PH.D., is an Assistant Professor in the School of Teacher Education at The University of Georgia. A former secondary physics and physical science teacher, he has developed interactive multimedia for science, mathematics, and teacher preparation. Dr. Hale received his Ph.D. in 1988 from The University of Georgia in science education. He is currently researching the use of multimedia technology to support teacher self-reflection and qualitative data analysis.

MICHAEL A. OREY, ED.D., is an associate professor of instructional technology at The University of Georgia where he teaches courses in Artificial Intelligence, Cognitive Science and Computer-Based Education. Since receiving his doctorate at Virginia Tech in 1989, he has developed several multimedia intelligent tutoring systems and conducted experimental evaluations of those systems. His current research interests include cognitive applications to learning, constructivism, and situated cognition.

THOMAS C. REEVES, Ph.D., is a professor of instructional technology at The University of Georgia where he teaches evaluation and multimedia design. Since receiving his Ph.D. at Syracuse University in 1979, he has developed and evaluated numerous interactive multimedia programs for education and training. He has been an invited speaker in Australia, Brazil, Bulgaria, Finland, Peru, Russia, South Africa, Switzerland, Taiwan, and the USA. His research interests include: 1) evaluation, 2) mental models, 3) electronic performance support systems, and 4) instructional technology in developing countries.

References

Blattner, M. M., & Dannenberg, R. B. (Eds.). (1992). *Multimedia interface design*. New York: Addison-Wesley.

Gery, G. (1991). *Electronic performance support systems*. Boston, MA: Weingarten.

Gomoll, K. (1990). Some techniques for observing users. In B. Laurel (Ed.), *The art of human-computer interface design* (pp. 85-90). New York: Addison-Wesley.

Hale, M., Gustafson, K., & Yeany, R. (1989, November). *Development and demonstration of science software tools for interactive video*. Paper presented at the annual meeting of the Association for the Development of Computer-based Instructional Systems, Crystal City, VA.

Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product & process*. New York: John Wiley & Sons.

Jih, H. J., & Reeves, T. C. (1992). Mental models: A research focus for interactive learning systems. *Educational Technology Research and Development, 40*(3), 39-53.

Kelly, J. (1994). *What do they think they're doing? The mental models of users of on-line public access catalogs*. Unpublished dissertation prospectus, The University of Georgia.

Laurel, B. (Ed.). (1990). *The art of human-computer interface design*. New York: Addison-Wesley.

Leiser, B. (1992). The presence phenomenon and other problems of applying mental models to user interface design and evaluation. In Y. Rogers, A. Rutherford, & P. A. Bibby (Eds.), *Models in the mind: Theory, perspective and application* (pp. 267-281). San Diego: Academic Press.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Polson, P. G. (1988). The consequences of consistent and inconsistent user interfaces. In R. Guindon (Ed.), *Cognitive science and its applications for human-computer interaction* (pp. 59-108). Hillsdale, NJ: Lawrence Erlbaum.

Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.

Tsao, P., Hale, M. & Fan, H. (1994, January). *Videotape Data Analyzer - Introducing an aid for helping analyze videotape data*. Qualitative Research in Education Conference, Athens, GA.